

BM



**PCT**  
WELTORGANISATION FÜR GEISTIGES EIGENTUM  
Internationales Büro  
INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE  
INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)

<p>(51) Internationale Patentklassifikation <sup>6</sup> : <b>G01N 33/68</b></p>	<b>A2</b>	<p>(11) Internationale Veröffentlichungsnummer: <b>WO 99/38013</b></p> <p>(43) Internationales Veröffentlichungsdatum: <b>29. Juli 1999 (29.07.99)</b></p>		
<table style="width: 100%; border: none;"> <tr> <td style="width: 45%; vertical-align: top; padding: 5px;"> <p>(21) Internationales Aktenzeichen: <b>PCT/DE99/00220</b></p> <p>(22) Internationales Anmeldedatum: <b>22. Januar 1999 (22.01.99)</b></p> <p>(30) Prioritätsdaten: <b>198 02 576.9      23. Januar 1998 (23.01.98)      DE</b></p> <p>(71) Anmelder (für alle Bestimmungsstaaten ausser US): <b>XERION PHARMACEUTICALS GMBH [-/DE]; Fraunhoferstrasse 9, D-82152 Martinsried (DE).</b></p> <p>(72) Erfinder; und (75) Erfinder/Anmelder (nur für US): <b>GE, Liming [-/DE]; Portiastrasse 12, D-81545 München (DE). ILAG, Leodevico [-/DE]; (DE). JOCELYN, H., Ng. [-/DE]; Knorrstrasse 85, D-80807 München (DE).</b></p> <p>(74) Anwalt: <b>SCHÜSSLER, Andrea; Truderinger Strasse 246, D-81825 München (DE).</b></p> </td> <td style="width: 55%; vertical-align: top; padding: 5px;"> <p>(81) Bestimmungsstaaten: <b>AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO Patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</b></p> <p><b>Veröffentlicht</b> <i>Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.</i></p> </td> </tr> </table>			<p>(21) Internationales Aktenzeichen: <b>PCT/DE99/00220</b></p> <p>(22) Internationales Anmeldedatum: <b>22. Januar 1999 (22.01.99)</b></p> <p>(30) Prioritätsdaten: <b>198 02 576.9      23. Januar 1998 (23.01.98)      DE</b></p> <p>(71) Anmelder (für alle Bestimmungsstaaten ausser US): <b>XERION PHARMACEUTICALS GMBH [-/DE]; Fraunhoferstrasse 9, D-82152 Martinsried (DE).</b></p> <p>(72) Erfinder; und (75) Erfinder/Anmelder (nur für US): <b>GE, Liming [-/DE]; Portiastrasse 12, D-81545 München (DE). ILAG, Leodevico [-/DE]; (DE). JOCELYN, H., Ng. [-/DE]; Knorrstrasse 85, D-80807 München (DE).</b></p> <p>(74) Anwalt: <b>SCHÜSSLER, Andrea; Truderinger Strasse 246, D-81825 München (DE).</b></p>	<p>(81) Bestimmungsstaaten: <b>AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO Patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</b></p> <p><b>Veröffentlicht</b> <i>Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.</i></p>
<p>(21) Internationales Aktenzeichen: <b>PCT/DE99/00220</b></p> <p>(22) Internationales Anmeldedatum: <b>22. Januar 1999 (22.01.99)</b></p> <p>(30) Prioritätsdaten: <b>198 02 576.9      23. Januar 1998 (23.01.98)      DE</b></p> <p>(71) Anmelder (für alle Bestimmungsstaaten ausser US): <b>XERION PHARMACEUTICALS GMBH [-/DE]; Fraunhoferstrasse 9, D-82152 Martinsried (DE).</b></p> <p>(72) Erfinder; und (75) Erfinder/Anmelder (nur für US): <b>GE, Liming [-/DE]; Portiastrasse 12, D-81545 München (DE). ILAG, Leodevico [-/DE]; (DE). JOCELYN, H., Ng. [-/DE]; Knorrstrasse 85, D-80807 München (DE).</b></p> <p>(74) Anwalt: <b>SCHÜSSLER, Andrea; Truderinger Strasse 246, D-81825 München (DE).</b></p>	<p>(81) Bestimmungsstaaten: <b>AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO Patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</b></p> <p><b>Veröffentlicht</b> <i>Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.</i></p>			
<p>(54) Title: <b>METHOD FOR SIMULTANEOUS IDENTIFICATION OF PROTEINS AND BONDING PARTNERS</b></p> <p>(54) Bezeichnung: <b>VERFAHREN ZUR GLEICHZEITIGEN IDENTIFIZIERUNG VON PROTEINEN UND IHREN BINDUNGSPARTNERN</b></p> <p>(57) Abstract</p> <p>The invention relates to a method for simultaneous identification of a protein and its binding partners, characterised in that a) proteins or aggregates of proteins from a biological source are isolated and separated, b) the separated proteins or aggregates of proteins are immobilized on a surface, c) a combinatory bank is incubated with proteins or aggregates of proteins immobilized on a surface, d) the members of the combinatory bank which bond with the immobilized proteins are separated from non-bonded members of said bank, e) the surface bonded complexes made of protein and bonding partners from the combinatory bank are isolated, f) the proteins in the complexes thus isolated are identified according to a mixed physical-chemical process and g) the isolated bonding partners are optionally enriched. The invention enables simultaneous identification of proteins with or without prior purification, and makes it possible to select members of combinatory banks which interact with said proteins. This allows the function of said proteins to be determined in a simple manner, using their specific bonding partners.</p> <p>(57) Zusammenfassung</p> <p>Die Erfindung betrifft Verfahren zur gleichzeitigen Identifizierung eines Proteins und seines Bindungspartners, das dadurch gekennzeichnet ist, daß man a) Proteine oder Proteinaggregate aus einer biologischen Quelle isoliert und auftrennt; b) die aufgetrennten Proteine oder Proteinaggregate auf einer Oberfläche immobilisiert; c) eine kombinatorische Bank mit den auf einer Oberfläche immobilisierten Proteinen oder Proteinaggregaten inkubiert; d) diejenigen Mitglieder der kombinatorischen Bank, die an die immobilisierten Proteine binden, von nicht-gebundenen Mitgliedern der Bank trennt; e) die an die Oberfläche gebundenen Komplexe aus Protein und Bindungspartner aus der kombinatorischen Bank isoliert; f) die Proteine in den so isolierten Komplexen mit einer Kombination aus einem physikalisch-chemischen Verfahren identifiziert; und g) gegebenenfalls die isolierten Bindungspartner anreichert. Die vorliegende Erfindung erlaubt somit die gleichzeitige Identifizierung von Proteinen mit und ohne vorherige Reinigung, sowie die Auswahl von Mitgliedern kombinatorischer Banken, die mit diesen Proteinen wechselwirken. Dadurch läßt sich auf einfache Weise die Funktion der Proteine über ihre spezifischen Bindungspartner ermitteln.</p>				

# **LEDIGLICH ZUR INFORMATION**

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidschan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland	ML	Mali	TR	Türkei
BG	Bulgarien	HU	Ungarn	MN	Mongolei	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MR	Mauretanien	UA	Ukraine
BR	Brasilien	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Island	MX	Mexiko	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	NE	Niger	UZ	Usbekistan
CF	Zentralafrikanische Republik	JP	Japan	NL	Niederlande	VN	Vietnam
CG	Kongo	KE	Kenia	NO	Norwegen	YU	Jugoslawien
CH	Schweiz	KG	Kirgisistan	NZ	Neuseeland	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	PL	Polen		
CM	Kamerun	KR	Republik Korea	PT	Portugal		
CN	China	KZ	Kasachstan	RO	Rumänien		
CU	Kuba	LC	St. Lucia	RU	Russische Föderation		
CZ	Tschechische Republik	LI	Liechtenstein	SD	Sudan		
DE	Deutschland	LK	Sri Lanka	SE	Schweden		
DK	Dänemark	LR	Liberia	SG	Singapur		
EE	Estland						

## **Verfahren zur gleichzeitigen Identifizierung von Proteinen und ihren Bindungspartnern**

Die Erfindung betrifft ein Verfahren zur gleichzeitigen Identifizierung von Proteinen und ihren spezifischen Bindungspartnern. Insbesondere betrifft die Erfindung ein Verfahren zur gleichzeitigen Identifizierung aller Proteine einer biologischen Quelle und ihrer entsprechenden Bindungspartner und somit die gleichzeitige Aufklärung von Struktur und Funktion aller Proteine aus einer biologischen Quelle.

Fortschritte in dem Humangenomprojekt führten zu einer nahezu unübersehbaren Menge von Daten aus den Genomen verschiedenster Organismen. Das vollständige Genom des Menschen wird vermutlich im Jahr 2003 sequenziert sein. Gegenwärtig ist das Genom von elf Mikroorganismen entschlüsselt. Eine genomische Sequenz erlaubt jedoch keine Aussage dahingehend, ob gegebene Proteine tatsächlich exprimiert werden, und wie sie im biologischen Gewebe funktionieren. Da die Proteine als die tatsächlichen funktionellen Gegenstücke der Gene den jeweiligen biologischen Zustand ihres Wirts bestimmen, spiegelt eine direkte Identifizierung der Proteine viel genauer den Zustand der biologischen Quellen des Wirts (i.e. des jeweiligen aktiven Genoms) wieder als die entsprechenden Gensequenzen.

Die Sequenzierung der mRNAs über cDNAs oder über Codesequenzen für die Expression (Expression Sequence Tags; ESTs) ergibt eine potentielle Korrelation

- 2 -

zwischen den in einer biologischen Quelle produzierten mRNAs und ihren Proteinäquivalenten. Dieses Vorgehen wird gegenwärtig am häufigsten angewendet. Jedoch entspricht bekanntlich die Genexpression nicht völlig der mRNA-Produktion. So enthalten ESTs RNA-Spleißintermediate, und der endgültige mRNA-Gehalt hängt entscheidend von der Stabilität der mRNAs ab. Aufgrund von Unterschieden in der Translationshöhe, der Stabilität, des Spleißmusters, posttranskriptioneller und posttranslationeller Modifikationen kann das Proteinendprodukt meist nicht aus den entsprechenden mRNAs oder ESTs vorhergesagt werden. Beispielsweise wurde ein Korrelationsfaktor von 0,43 (i.e. keine Entsprechung im Verhältnis eins zu eins) zwischen der gebildeten mRNA und der tatsächlichen Menge an GST- $\pi$ -Protein, das in den verschiedenen Geweben exprimiert wird (Anderson, L., IBC's International Conference on proteomics, Boston, MA, 1997). Auf ähnliche Weise fand sich keine Korrelation zwischen der mRNA-Produktion und dem Vorkommen der Proteine unter 23 in der menschlichen Leber gebildeten Proteinen (Large Scale Biology)

Um die Funktion unbekannter Proteine zu identifizieren wurden verschiedene Methoden angewendet. Beispiele hierfür sind der Vergleich unbekannter Proteine mit sequenzhomologen Proteinen ähnlicher Funktion, Antisense-Technik, Knock-out-Tiermodelle oder der Einsatz transgener Tiere. Jedoch läßt sich mit diesen Methoden die Funktion interessierender Proteine nicht direkt identifizieren (z.B. beim homologen Vergleich), oder das Verfahren ist aufwendig, zeitraubend und damit wenig geeignet für ein Massenscreening (z.B. bei Knock-out oder transgenen Tieren).

- 3 -

Proteinproben lassen sich mittels 2D-Gelelektrophorese auftrennen, und die Proteine können mittels Massenspektrometrie in Kombination mit EST- und Protein-Datenbankrecherchen identifiziert und bestätigt werden. Je mehr ESTs ermittelt werden, desto leichter wird es möglich, das Protein in voller Länge nur über Peptidfragmente zu identifizieren. Jedoch führt die Identifikation der Proteine und ihrer Modifikationen nicht zur Ermittlung ihrer Funktion.

Die Funktion einiger Proteine kann über ihre Wechselwirkung mit anderen Proteinen, deren Funktion bekannt ist, bestimmt werden. Methoden wie die Hefe-zwei-Hybrid-Technik (Phizicky, E.M. und Fields, S., Microbiological Rev., 59:94-123, 1995) und die Phagendisplay-Technik (Ilag, V. und Ge, L., PCT/EP97/00931, 1996) für ein wechselseitiges Absuchen von Genbanken können angewendet werden, um derartige Wechselwirkungen zu ermitteln. Beispielsweise bietet das Hefe-zwei-Hybrid-Verfahren die Möglichkeit, Proteine und ihre Wechselwirkungen in einem eukaryotischen Wirt zu untersuchen; das Verfahren besitzt jedoch den Nachteil, daß die Wechselwirkung nur im Hefezellkern stattfindet und nur auf nicht-sezernierte Proteine angewendet werden kann. Obwohl das Phagen-Display-Verfahren diese Probleme überwindet, ist nicht zu erwarten, daß die Proteine in ihrer nativen Form vorliegen, da möglicherweise posttranslationelle Modifikationen stattfinden.

Die am häufigsten verwendete funktionelle Analyse beruht auf immunologischen Methoden, z.B. histochemische Analyse, FACS oder Immunpräzipitation. Wegen ihrer hohen Spezifität und Affinität werden am häufigsten Antikörper eingesetzt. So können Liganden-Rezeptor-Wechselwirkungen durch blockierende Antikörper

blockiert werden. Jedoch können unter Verwendung herkömmlicher Methoden Antikörper nur gegen fremde Antigene erzeugt werden. Beispielsweise ist es schwierig oder sogar unmöglich, menschliche Antikörper zu erzeugen, die in auf Antikörpern beruhenden Therapien gegen Antigene menschlichen Ursprungs nützlich sind. Antikörper sind jedoch in der Herstellung und Reinigung teuer und aufwendig.

Um die entsprechenden Bindungspartner für ein Antigen zu ermitteln, werden Proteinbanken, insbesondere Antikörperbanken (einkettiges Fv/scFV und Fab), und Peptidbanken funktionell auf der Oberfläche von filamentösen Bakteriophagenteilchen in einem Phagen-Display-System exprimiert (vgl. Smith, G.P., *Science*, 228: 1315-1317, 1985). Die scFvs, Fabs, Proteine oder Peptide werden an eine Komponente eines Oberflächenproteins des Phagen fusioniert, wodurch die Bindung des interessierenden scFvs, Fabs, Proteins oder Peptids stattfinden kann. Die Gene, die die präsentierten Proteine codieren, werden in die Phagenteilchen verpackt, und so binden die Proteinprodukte direkt an ihre genetische Information. Menschliche scFv-Phagenbanken wurden zur Isolierung therapeutisch wichtiger Antikörper vielfach verwendet (vgl. Vaughan, T.J. et al., *Nature Biotechnol.* 14:309-314, 1996).

Es konnte auch gezeigt werden, daß Antikörper mit hoher Affinität und Spezifität gegen Selbstantigene aus den menschlichen kombinatorischen Antikörperbank isoliert werden können (Griffiths A.D. et al., *EMBL J.*, 12:725-734, 1993). Jedoch sind bei all diesen Technologien gereinigte Proteine in beträchtlicher Menge erforderlich, wodurch sie nur auf die in größter Menge vorkommenden

Proteine oder Proteine, die rekombinant hergestellt werden können, angewendet werden können. Obwohl es möglich ist, Antikörper oder Peptide gegen Proteine auf der Zelloberfläche zu richten, muß die Identität dieser Proteine bekannt sein.

Die WO 94/26787 und die WO 97/22972 beschreiben die Isolierung von Antikörpern aus kombinatorischen Antikörperbanken gegen nicht gereinigte und zuvor nicht identifizierte Zelloberflächenantigene bzw. intrazelluläre krankheitsspezifische Antigene. Nach dem Verfahren gemäß der WO 94/26787 können jedoch nur Zelloberflächenantigene ermittelt werden; die unbekannten Antigene können nicht direkt identifiziert werden. Bei dem Verfahren gemäß der WO 97/22972 müssen zuerst antigenspezifische Antikörper aus einem Selektionsprozeß verfügbar gemacht werden. Daher müssen in einer ersten Stufe antigenspezifische Antikörper selektioniert werden, bevor die unbekannten Antigene identifiziert werden können.

Es besteht somit ein Bedarf nach einem Verfahren, alle in einer ausgewählten biologischen Quelle produzierten Proteine direkt abtrennen und funktionell identifizieren zu können. Mit einem derartigen Verfahren ließen sich alle Proteine, die den Phänotyp der biologischen Quelle, z.B. Gewebe, Mikroorganismen, Zellkulturen etc. bestimmen, direkt identifizieren und funktionell charakterisieren.

Der vorliegenden Erfindung liegt daher die Aufgabe zugrunde, ein Verfahren bereitzustellen, mit dem gleichzeitig Proteine und ihre spezifischen Bindungspartner aus einer kombinatorischen Bank isoliert werden können, ohne daß eine

- 6 -

vorherige Trennung, Reinigung und Identifizierung der Proteine erforderlich sind. Mit dem Verfahren sollen alle Proteine aus einer biologischen Quelle funktionell erfaßt werden können. Es soll also der Proteomstatus einer biologischen Probe ermittelt werden können. Das Verfahren soll auf alle möglichen Bindungspartner eines Proteins anwendbar sein. Ferner soll das Verfahren zusätzlich die Identifizierung der funktionell charakterisierten Proteine erlauben. Das Verfahren soll einfach, schnell und kostengünstig durchzuführen sein und sich für ein Massenscreening und zur Automatisierung eignen. Ferner soll das erfindungsgemäße Verfahren eine rasche, eindeutige und einfache Diagnostik von Stoffwechselkrankheiten sowie die Identifizierung von Arzneimittelwirkungen erlauben. Mit dem erfindungsgemäßen Verfahren soll eine Datenbank aus Proteinen und ihrem jeweiligen spezifischem Bindungspartner erstellt werden können. Diese Datenbank soll die Ermittlung von Arzneimittelprototypen unterstützen.

Diese Aufgabe wird erfindungsgemäß durch ein Verfahren zur gleichzeitigen Identifizierung eines Proteins und seines Bindungspartners gelöst, das dadurch gekennzeichnet ist, daß man

- a) Proteine oder Proteinaggregate aus einer biologischen Quelle isoliert und auftrennt,
- b) die aufgetrennten Proteine oder Proteinaggregate auf einer Oberfläche immobilisiert,
- c) eine kombinatorische Bank mit den auf einer Oberfläche immobilisierten Proteinen oder Proteinaggregaten inkubiert,
- d) diejenigen Mitglieder der kombinatorischen Bank, die an die immobilisierten Proteine binden, von nicht-gebundenen Mitgliedern der Bank trennt,



- 7 -

- e) die an die Oberfläche gebundenen Komplexe aus Protein und Bindungspartner aus der kombinatorischen Bank isoliert,
- f) die Proteine in den so isolierten Komplexen mit einer Kombination aus einem physikalisch-chemischen Verfahren identifiziert, und
- g) gegebenenfalls die isolierten Bindungspartner anreichert.

Bevorzugt wird das Verfahren derart durchgeführt, daß man bei Schritt a) die biologische Probe in einem geeigneten Puffer solubilisiert und die Probe unter Verwendung eines Proteintrennverfahrens ausgewählt aus 2D-Gelelektrophorese, Perfusionschromatographie, Flüssigchromatographie oder Kapillarelektrophorese auftrennt. Weiter bevorzugt werden die aufgetrennten Proteine auf den Kavitäten einer Mikrotiterplatte immobilisiert oder auf eine Membran geblottet oder mit mit spezifischen Antikörpern beschichteten Mikrokügelchen eingefangen.

Der Ausdruck Proteinaggregate bezeichnet einen Zusammenschluß mehrerer Proteine. Dieser Zusammenschluß kann das Ergebnis einer funktionellen Assoziation mehrerer Proteine sein, z.B. Enzyme des Krebs-Zyklus, die zu einer funktionellen Einheit verbunden sind, an der die Umwandlung des Substrats in das Endprodukt über mehrere Zwischenstufen erfolgt. Der Zusammenschluß kann aber auch das Ergebnis des Trennverfahrens sein, und umfaßt z.B. in einem chromatographischen Peak eluierende Proteine, die durch eine gemeinsame Retentionszeit verbunden sind.

Bevorzugt verwendet man als kombinatorische Bank Random-Peptid-Banken,

(scFv)Banken der Immunglobulinsuperfamilie, Protein-Display-Banken, kombinatorische chemische Banken, RNA- oder DNA-Banken. Bevorzugt wird Schritt g) so durchgeführt, daß man einen bakteriellen Wirt mit den isolierten Proteinen infiziert, um ausgewählte Phagenteilchen zu vermehren, und die ausgewählten Proteinbindungspartner sequenziert, oder die ausgewählten Bindungspartner durch einzigartige Sequenzanknüpfungen (tag) identifiziert.

Überraschenderweise wurde gefunden, daß Proteine und ihre Bindungspartner gleichzeitig identifiziert werden können, wenn man ohne vorherige Reinigung und Identifizierung Proteine nach Auftrennung mit einer kombinatorischen Bank inkubiert und die so erhaltenen Komplexe aus Protein und Bindungspartner einem physikalisch-chemischen Identifizierungsverfahren unterwirft.

Da sich erfindungsgemäß die Proteinproben direkt von ihren biologischen Quellen ableiten, wird keine teure, zeitaufwendige oder risikobehaftete Proteinproduktion benötigt. Da der Selektionsschritt in einer Stufe erfolgt, werden keine Anreicherung der Bank oder eine anschließende Selektion benötigt. Das erfindungsgemäße Verfahren eignet sich daher ideal zur Produktion von proteinspezifischen Antikörpern oder Peptiden in großem Umfang. Ferner kann jede kombinatorische Bank eingesetzt werden.

Da die aufgetrennten Proteine/Proteinkomplexe direkt aus dem Gel oder in immobilisierter Form unter Verwendung von Massenspektrometrie identifiziert werden können, kann das Verfahren der Proteinidentifizierung und Erzeugung proteinspezifischer Subbanken als paralleles Verfahren durchgeführt werden.

Dadurch eignete sich das erfindungsgemäße Verfahren für Proteinidentifizierungen in großem Umfang.

Das erfindungsgemäße Verfahren kann mit jeder beliebigen kombinatorischen Bank durchgeführt werden, z.B. Proteinbank, Peptidbank, cDNA-Bank, mRNA-Bank, Bank mit organischen Molekülen, scFv-Bank mit Immunglobulinsuperfamilie, Proteindisplay-Bank etc.. In den Banken können präsentiert sein: alle Arten von Proteinen, z.B. Strukturproteine, Enzyme, Rezeptoren, Liganden, alle Arten von Peptiden einschließlich Modifikationen, DNAs, RNAs, Kombinationen von DNAs und RNAs, Hybride von Peptiden und RNA oder DNA, alle Arten von organischen Molekülen, z.B. Steroide, Alkaloide, Naturstoffe, synthetische Stoffe etc. Die Präsentation kann auf verschiedene Arten erfolgen, z.B. als Phagen-Display-System (z.B. filamentöse Phagen wie M13, fl,fd etc., lambda-Phagen-Display, virales Display etc.), Präsentation auf Bakterienoberflächen, Ribosomen etc.

Die kombinatorische Bank kann hergestellt werden durch:

- a) Konstruktion von Random-Peptid-Banken, in denen Banken präsentiert werden können,
- b) Konstruktion von scFv-Banken oder Banken von beliebigen Mitgliedern der Immunglobulin-Superfamilie, in denen Mitglieder der Banken präsentiert werden können,
- c) Konstruktion von Proteinbanken, in denen Proteine präsentiert werden können,
- d) Konstruktion von kombinatorischen chemischen Banken, in denen organi-

- 10 -

- sche Moleküle der Banken präsentiert werden können,
- e) Konstruktion von RNA- oder DNA-Banken in denen die ausgewählten Mitglieder der Banken isoliert und über geeignete Oligoprimer amplifiziert werden können.

Derartige Verfahren sind einem Fachmann auf dem Gebiet bekannt.

Die erfindungsgemäß zu identifizierenden Proteine können aus jeder biologischen Quelle stammen, z.B. aus gesunden oder erkrankten Geweben, Zellkulturen, Organpräparaten, Körperflüssigkeiten, Biopsieproben aller Art, Organkulturen, Mikroorganismen, Pflanzenpräparate, etc.

Die vorliegende Erfindung erlaubt somit die gleichzeitige Identifizierung von Proteinen mit und ohne vorherige Reinigung, sowie die Auswahl von Mitgliedern kombinatorischer Banken, die mit diesen Proteinen wechselwirken. Dadurch läßt sich auf einfache Weise die Funktion der Proteine über ihre spezifischen Bindungspartner ermitteln.

Die Identifizierung von Proteinen einer Expressionsfamilie ergänzt oder ersetzt sogar das Verfahren zur Identifizierung von Genen: Die Identifizierung von Proteinen in ihrem nativen Zustand bestätigt die entsprechenden Gensequenzen oder identifiziert mögliche posttranskriptionelle und posttranslationelle Modifikationen. Die Identifizierung der meisten oder aller Proteine aus einer nicht sequenzierten oder teilweise sequenzierten biologischen Probe beschleunigt Bemühungen zur Ermittlung einer Genzielsequenz oder vermeidet den aufwendigen Prozeß der Gensequenzierung. Die Information auf der Proteinebene spiegelt die biologi-

sche Identität eines Organismus besser wieder als die Information auf genomischer Ebene. So spiegelt die Identifizierung aller Proteine einer biologischen Probe die relevante Information über den biologischen Zustand der Probe und damit des untersuchten Organismus bzw. Teil des Organismus wieder. Ferner ist zu erwarten, daß die Variation im Vorkommen eines Proteins dynamisch ist, i.e. sie wird von endogenen und exogenen Faktoren beeinflußt. Es ist nicht möglich, diese Zusatzinformation aus der Sequenzierung des Genoms oder der mRNA abzuleiten (s.o.).

Ferner können erfindungsgemäß spezifisch miteinander wechselwirkende Mitglieder einer kombinatorischen Bank gegen die meisten oder gar alle Proteine einer gegebenen biologischen Probe gleichzeitig ermittelt werden. Dies ersetzt das aufwendige Verfahren der Genisolierung, Subklonierung, Expression und Reinigung des rekombinanten Proteins. Ferner ist es mit den gegenwärtigen bekannten Methoden nicht nur aufwendig und schwierig, spezielle Mitglieder einer kombinatorischen Bank gegen ein Protein zu isolieren, sondern es ist auch unsicher, ob alle Proteine exprimiert werden können und ob die exprimierten Proteine tatsächlich in ihrer nativen Form vorliegen. Beispielsweise ist gut bekannt, daß das Glycosylierungsmuster von Proteinen eukaryotischen Ursprungs entscheidend von den zur Expression gewählten Wirten abhängt. Ferner ist es praktisch nicht möglich, spezifisch wechselwirkende Mitglieder einer Bank gegen die meisten oder alle Proteine aus komplizierten biologischen Proben gleichzeitig unter Verwendung bisher bekannter Methoden zu erhalten.

Diese Kombinationen aus Protein und spezifischem Mitglied einer kombinatori-

schen Bank kann verwendet werden, um den Proteomstatus der Probe und nicht nur von einigen wenigen Proteinen gleichzeitig zu überwachen. Dieser Fortschritt in der Proteomüberwachung soll die Ermittlung von Angriffsorten von Arzneimitteln und von pharmazeutischen Leitsubstanzen zur Weiterentwicklung zu Arzneimitteln beschleunigen, da die meisten Erkrankungen, die gegenwärtig therapiert werden, multifaktorieller Genese sind, d.h. mehr als ein Protein an ihrer Genese beteiligt ist.

Ferner ermöglicht die vorliegende Erfindung die Anwendung jeder kombinatorischen Bank, z.B. Protein-, Peptid- oder Antikörperbanken oder DNA- und RNA-Banken oder sogar mit einer Codesequenz versehene synthetische organische Molekülbanken (Brenner, S. und Lerner, R.A., Proc. Natl. Acad. Sci. USA, 89,5381-5383, 1992). Die Erfindung besitzt somit beispielsweise die folgenden technischen Anwendungsmöglichkeiten:

- a) Isolierung von Proteinen, die für Proteine aus einer biologischen Probe spezifisch sind, und Erstellung einer Protein-Protein-Wechselwirkungsbank (z.B. Netzwerk von signalübertragenden Stoffwechselwegen, wenn Proteine der kombinatorischen Bank von der gleichen biologischen Quelle stammen wie die Probe) oder einer Krankheitserreger-Wirt-Datenbank (z.B. wenn Proteine der kombinatorischen Bank dem Krankheitserreger angehören und die Probe der Wirt ist),
- b) Isolierung und Identifizierung proteinspezifischer Peptide, mit dem Ziel der Ermittlung potentieller Peptid-agonisten oder -antagonisten oder peptidomimetischer Moleküle zum Design von neuen Arzneimitteln,
- c) Isolierung und Identifizierung rekombinanter Antikörper mit den vorste-

- hend genannten Anwendungsspektren,
- d) Isolierung und Identifizierung proteinspezifischer DNA- oder RNA-Moleküle, die für Anwendungen von Protein-Knock-out zur Funktionsermittlung bis zur Arzneimittel-entwicklung geeignet sind, und
  - e) Isolierung und Identifizierung kleiner organischer Moleküle, die direkte Moleküle für die Entwicklung von Anzneimitteln sind.

Es ist klar, daß das erfindungsgemäße Verfahren nicht nur das Verständnis biologische Prozesse beschleunigt, sondern auch die Aufklärung von Krankheitsmechanismen und die gezielte Entwicklung neuer Arzneimittel beschleunigt.

Ferner erlaubt das erfindungsgemäße Verfahren eine einfache und rasche Diagnose von Stoffwechselerkrankungen. Gegenwärtig wird Diagnostik nur an einem speziellen Protein oder Stoffwechselprodukt durchgeführt. Bezüglich Erbkrankheiten werden nur ganz spezielle Mutationen in einem Gen getestet, jedoch mit unsicherer Aussagekraft. Obwohl es möglich ist, eine Diagnostik mit mehreren Proteinen oder Metaboliten durchzuführen, sind derartige Verfahren teuer und zeitaufwendig und erlauben nur eine begrenzte Information über die Krankheit. Die Überwachung des Proteomstatus einer biologischen Probe erlaubt - wie vorstehend ausgeführt - die Feststellung des aktuellen Zustands in einem biologischen Gewebe. Mit dem erfindungsgemäßen Verfahren kann nun der Proteomstatus eines Gewebes einfach, schnell und zuverlässig ermittelt werden. Dazu wird nach dem erfindungsgemäßen Verfahren eine proteinspezifische Bank oder Subbank eines Gewebes, z.B. der Leber, der Lunge, des Magens etc. erzeugt und gegebenenfalls vielfach amplifiziert. Eine derartige Bank wird

dann auf einer geeigneten Oberfläche, z.B. aus Glas, Kunststoff, einem Halbleiterchip einer optischen Faser oder einer CD in an sich bekannter Weise immobilisiert. Die Bindung der Proteine an ihre verwandten Subbanken, die auf der geeigneten Oberfläche immobilisiert sind, können unter Verwendung bereits bekannter Nachweisverfahren, nachgewiesen werden z.B. BIAcore-Chips. Umgekehrt kann auch die biologische Probe auf der Oberfläche eines derartigen Chips immobilisiert werden. Die Oberfläche selbst wird bevorzugt in mehrere Mikrokompartimente unterteilt, wobei jede Unterteilung einem speziellen Protein zugeordnet wird, dessen kombinatorische Subbank zuvor erzeugt worden ist. Nach Abwaschen überschüssiger nicht gebundener Probenlösung werden die proteinspezifischen Antikörper- oder Proteinsubbanken jedem Kompartiment zugesetzt. Die gebundenen Antikörper oder Peptide können durch ihre fusionierten Peptid-Tags identifiziert werden. Aus den so erhaltenen Signalen (qualitativ/quantitativ) kann dann durch Vergleich mit dem Signalmuster einer normalen Probe auf eine entsprechende Erkrankung des Gewebes bzw. Organs geschlossen werden. Derartige Diagnosechips können zur Diagnose von Organ- und Gewebszuständen, Infektionen und Krankheiten aller Art gezielt angefertigt werden. Ferner kann ein solcher Chip auch zur Untersuchung von Stoffwechselzuständen in Pflanzen, Mikroorganismen etc. verwendet werden. So lassen sich mit einem solchen Chip gezielt und rasch Zustände wie ein frischer Herzinfarkt, ein Magengeschwür, eine Gewebsnekrose, Infektionen wie z.B. Hepatitis, Tropenkrankheiten, AIDS, Autoimmunerkrankungen aller Art etc. diagnostizieren.

Ein derartiger Diagnosechip kann in Form eines gebrauchsfertigen Kits angeboten werden. Ein derartiges Kit umfaßt einen entsprechenden gewebs- bzw. organ-



spezifischen Chip, auf den eine entsprechende Proteinbank immobilisiert ist, eine Ausrüstung zur Probenahme, z.B. Spritze, Skalpell etc. sowie ein Gefäß zur Durchführung der Inkubation zwischen Chip und Probe und Anleitungen zur Durchführung und Auswertung. Entsprechende gebrauchsfertige Kits können gezielt hergestellt werden.

Bei dem erfindungsgemäßen Verfahren wird zuerst die Proteinprobe aus der jeweiligen biologischen Quelle solubilisiert und aufgetrennt. Hierzu kann jedes auf dem Gebiet der Proteinabtrennung einem Fachmann bekannte Verfahren verwendet werden. Bevorzugt ist die 2D-Gelelektrophorese. Die so aufgetrennte Proteinprobe wird anschließend auf eine Oberfläche, bevorzugt eine Membran, geblottet. Es ist nicht notwendig, die Proteine in dieser Stufe zu identifizieren.

Als nächster Schritt wird eine kombinatorische Bank z.B. aus einem Peptid oder antikörperartigen Molekülen oder Proteinen, die auf der Oberfläche von filamentösen Phagenteilchen exprimiert sind, mit der Membran, auf die die aufgetrennte Proteinprobe geblottet ist, inkubiert. Die Inkubation kann nach einem Fachmann bekannten Bedingungen durchgeführt werden. Die kombinatorische Bank kann z.B. durch kombinatorische chemische Methoden wie randomisierte Oligokassetten, hergestellt werden und aus jeder biologischen Quelle (z.B. cDNA oder Antikörper aus immunisierten Tieren) isoliert werden. Die Inkubationsbedingungen werden dabei so gewählt, daß ein Teil der Proteine in der Bank an die einzelnen Proteine in der aufgetrennten Probe bindet. Die speziell auf den jeweiligen Fall anzuwendenden Bedingungen können von einem Fachmann durch einfache Routineversuche ermittelt werden.

Gemäß einer weiteren Ausführungsform werden Proteine, Peptide oder antikörperartige Moleküle, die in einem in vitro in einem Polysom-Display-System (Mattheakis, L.C. et al., Proc. Natl. Acad. Sci. USA, 91, 9022-9026, 1994) präsentiert sind, oder DNA- oder RNA-Molekülbanken, die mittels SELEX oder ähnlichen Systemen (Tuerk, C., und Gold, L. Science, 249: 505-510, 1990) erzeugt wurden, oder kombinatorische Banken organischer Moleküle mit den getrennten immobilisierten Molekülen inkubiert. Die Inkubation wird so ausgeführt, daß ein Teil der Mitglieder der Bank an die einzelnen getrennten Proteine in der Probe bindet.

Anschließend wird die Membran gründlich gewaschen, um zu gewährleisten, daß nur die Mitglieder der Bank, die für die getrennten Proteine spezifisch sind, haften bleiben. Die Waschbedingungen richten sich nach den jeweils gebundenen Proteinen und der verwendeten Bank. Die Flecken, die zu identifizierenden Proteinen entsprechen, werden ausgeschnitten und die an die Proteine gebundenen filamentösen Phagenteilchen werden eluiert. Die ausgewählten Phagenteilchen werden entweder aufbewahrt oder zur Infektion relevanter Wirtszellen verwendet, um die ausgewählten Mitglieder der Bank zu vermehren. Alternativ kann die genetische Information, die den gebundenen Mitgliedern der Bank entspricht, direkt mittels PCR gewonnen oder über ihre einzigartige Codesequenz identifiziert werden.

Gemäß einer alternativen Ausführungsform wird die geblottete Membran reversibel mit kolorimetrischen oder lumineszierenden Farbstoffen angefärbt und die gefärbten Proteinflecken werden automatisch analysiert. Die Farbe wird entfernt

- 17 -

und die Membran wird verwendet, um die präsentierte Bank durchzumustern. Die nun nicht mehr sichtbaren Proteinflecken werden automatisch registriert und ausgeschnitten.

Gemäß einer weiteren Ausführungsform der Erfindung werden die auf die Membran geblotteten Proteine nach einem beliebigen, einem Fachmann bekannten physikalisch-chemischen Verfahren zur Proteinidentifizierung identifiziert.

Bevorzugt wird erfindungsgemäß die Identität der interessierenden Proteine massenspektrometrisch (wie z.B. in Siuzdak, G., Mass Spectrometry for Biotechnology, Academic Press, Inc., 1996 beschrieben) mit anschließender Protein- oder EST-Datenbankrecherche (z.B. nach Mann, M., in Microcharacterization of Proteins, Hrsg. R. Kellner, F. Lottspeich, H.E. Meyer, VCH Weinheim, 1994) identifiziert. Die Proteine werden anschließend im Gel enzymatisch oder chemisch gespalten. Der Spaltansatz kann dann als solcher vollständig massenspektrometrisch analysiert werden oder vorher mittels Mikrosäulen-Flüssigchromatographie (LC) analysiert werden. Die massenspektrometrische Analyse kann auf verschiedene, an sich bekannte Arten durchgeführt werden, z.B. mit einer Ionisierungsquelle wie einem Elektrospray (Chapman, J.R., et al., Methods in Molecular Biology, 61, JR Chapman Hrsg., Humana Press Inv. Totowa NJ, USA, 1996) einschließlich Nanoelektrospray (Wilm. M. und Mann, M., Anal. Chem. 68, 1-8, 1996) und matrixunterstützter Laserdesorption und Ionisierung (MALDI) (Siuzdak, G. Mass Spectrometry for Biotechnology, Academic Press Inc. 1996) oder eine Kombination aus Massenanalysatoren wie Triple, Quadrupol, Flugzeit, Magnetsektor, Fourier-Transformations-Ionenzyklotron-Resonanz und

**Quadropol-Ioneneinfang.**

Wenn die Peptide aus dem Spaltansatz nicht ausreichen, die Identität des Proteins eindeutig aus der Datenbank zu identifizieren, kann durch eine weitere Fragmentierung im Massenspektrometer wie z.B. durch Zerfall nach der Quelle in MALDI-TOF, MS/MS (Tandem-Massenspektrometrie), MS<sup>n</sup> eine weitere Sequenzinformation für die Datenbankrecherche erhalten werden. Die Ergebnisse der Recherche werden dann durch Identifizieren der für die Recherche nicht verwendeten Peptidfragmente im Massenspektrum bestätigt.

Alternativ können die Proteine oder Proteinfamilien durch de novo-Sequenzierung (z.B. nach Shevchenko, A., et al., Rapid Communications in Mass Spectrometry, 11, 1015-1024, 1997) identifiziert werden. Dazu stehen beispielsweise die folgenden Methoden zur Verfügung:

- 1) Verwendung der Mikrosäulen-LC zur Auftrennung der Peptide im Spaltansatz (Dongre, A., et al., TIBTECH, 15, 418-425, 1997), gefolgt von einer automatischen Datengewinnung basierend auf vorgewählten Bedingungen und einem Computer-Algorithmus, um die Datenbank abzusuchen, wobei eine Korrelationsanalyse verwendet wird, um die Sequenzen der Datenbank dem Ergebnis der Tandem-Massenspektroskopieanalyse anzupassen. Das verwendete Massenspektrometer besitzt eine Elektrospray-Ionisationsquelle.
- 2) Analyse des kompletten Spaltansatzes (Shevchenko, A., et al., Proc. Natl. Acad. Sci. USA, 93, 14440-14445, 1996) mittels MALDI mit verzögerter

Extraktion und Automatisierung unter Verwendung eines Realzeit-Fuzzy-logic-Algorithmus, um die Massenspektren zu erhalten, und eines Software-Bindeglieds zu einer automatischen Datenbankrecherche. Proteine, die hier nicht identifiziert werden, werden dann einer Nanoelektrospray-Tandem-Massenspektrometrie mit Stammionenscanning (Wilm M. et al., Anal. Chem., 68, 527-533, 1996) unterworfen. Sequenzmarker (Mann, M., TIBS, 21, 494-495, 1996) werden dann zum Absuchen der Datenbank verwendet.

Eine Datenbank über die Protein-Protein-Wechselwirkung oder die Liganden-Zielmolekül-Wechselwirkung kann aufgrund der identifizierten Proteine oder Proteinkomplexe und ihrer verwandten Bindungspartner erstellt werden und ist ebenfalls Gegenstand der Erfindung.

Die Figur 1 zeigt in einem Fließschema des erfindungsgemäßen Verfahrens im Überblick. Die Identifizierung der getrennten Proteine kann entweder in der Immobilisierungs- oder der Trennstufe erfolgen.

Die folgenden Beispiele erläutern die Erfindung näher ohne sie zu beschränken.

**Beispiel 1****A. Identifizierung der Proteine aus den Mitochondrien**

Mehr als 100 mitochondriale Erkrankungen sind bekannt. Einige dieser Erkrankungen treten im Zusammenhang mit Alterungsprozessen oder neurologischen Prozessen auf. Beispiele hierfür sind Herzversagen, Demenz oder Schizophrenie. Einige dieser Erkrankungen werden durch Mutationen der mitochondrialen oder nukleären DNA, die die meisten der Struktur- und Regulatorproteine der Mitochondrien codiert, verursacht. Es ist somit wichtig, alle mitochondrialen Proteine zu identifizieren und ihre Expression sowohl in gesunden als auch erkrankten Geweben zu überwachen.

Im vorliegenden Beispiel wurden Mitochondrien aus gesunden Rinderherzen untersucht.

Die Mitochondrien aus Rinderherz wurden gemäß Smith, A.L., Methods Enzymol. 10, 81-86, 1967, präpariert. Submitochondriale Teilchen wurden daraus gemäß Cattell et al., Biochem J., 125, 169-177, 1971 hergestellt. Die Proteine wurden aus den Mitochondrien und submitochondrialen Teilchen nach dem Chloroform/Methanol-Extraktionsverfahren gemäß Fearnley I., und Walker, J.E., Biochem 26, 8247-8251, 1987 extrahiert.

Der Chloroform/Methanolextrakt enthielt etwa 15 Proteine. Die verschiedenen Proteine wurden durch Gelfiltration über Toyopearl HW-55 in einem Chloroform-/Methanol/Wassergemisch (46:46:8, bezogen auf das Volumen) mit 60 mM

Ammoniumacetat, pH 7 extrahiert.

Die aufgetrennte Proteinprobe wurde in einem Lösungsmittel aus Chloroform/-Methanol/wäßrige Ameisensäure (4:4:1, bezogen auf das Volumen) aufgelöst. Die Proben wurden direkt in ein Massenspektrometer über einen kontinuierlichen Fließträger aus dem gleichen Lösungsmittel oder unter Verwendung der Nanoelektrospray-Technik (Wilm, M.S., und Mann, M., Int. J. Mass Spectrom. Ion Processes, 136, 167-180, 1994) injiziert. Bei der zuletzt genannten Technik wurden etwa 2 µl der Proteinlösung in eine goldplattierte, zu einer etwa 1µm Düse ausgezogene Kapillare überführt. Ein Hauptvorteil dieses Verfahrens ist der geringere Materialbedarf, verglichen mit herkömmlicheren Ionisierungsmethoden. Die Massenspektren wurden aufgezeichnet, und entsprechende Molekulargewichte wurden berechnet.

Die Identität der Proteine kann mittels Tandem-Massenspektrometrie bestimmt werden. Aminosäureteilsequenzen wurden mit Tandem-MS mehrerer intakter Proteolipidionen durch Fragmentierung der Molekülionen im Massenspektrometer durch Kollision mit Argongas bestimmt. Aus den Massenunterschieden zwischen benachbarten Ionen in der Reihe der Peaks in dem Massenspektrum wurde eine Spanne der Aminosäuresequenz erhalten. Die so erhaltene Sequenz wurde dann verwendet, um die Sequenzen der SWISSPROT-Proteinbank durchzumustern, wobei das Programm PEPTIDE SEARCH (Mann, M. et al., Biol. Mass Spec., 22, 338-345, 1993) verwendet wurde, und die Identität mit dem Programm MACPROMASS (Lee, T.D., und Vemuri, S., Biomed. Environ. Mass Spectrom., 19, 639-645, 1990) bestätigt wurde.

Alternativ wurden die Proteine einzeln mit Trypsin gemäß Shevchenko et al., Proc. Natl. Acad. Sci. USA 93: 14440-14444, 1996 gespalten. Ein Aliquot des Überstands wurde entnommen und mittels MALDI-Peptidmapping analysiert (Shevchenko, a.a.O.). Das Programm PEPTIDE SEARCH wurde verwendet, um die Peptidmassenkartierung des isolierten Proteins zu vergleichen. In Fällen, in denen die Peptidmassenkartierung zu keiner eindeutigen Identifizierung führte, wurden die Proben mittels Nanoelektrospray-Massenspektrometrie untersucht. Das Peptidgemisch wurde auf einer Kapillare aus 50 nl Poros R2-Harz (PerSeptive Biosystem, Framingham, MA) mikrogereinigt. Die Peptide wurden gewaschen und dann in einem Stufengradienten mit 0,5 µl 50% Methanol in 5% Ameisensäure in eine Nanoelektrospraykapillare eluiert. Diese Kapillare wurde in ein Massenspektrometer überführt und die Probe wurde etwa 20 min versprüht. In dieser Zeitspanne wurden aus dem Massenspektrum hervorgehende Peptidionen ausgewählt, isoliert und in der Kollisionskammer des Massenspektrometers fragmentiert. Aus den Tandem-Massenspektren wurden kurze Sequenzstücke zu Peptidsequenztags aneinandergefügt und mit einer Proteinsequenzdatenbank oder einer EST-Datenbank unter Verwendung von PEPTIDE SEARCH verglichen.

Aus dem Enzympool in Rindermitochondrien kann Cytochromoxidase mit acht Untereinheiten (36, 21, 19, 14, 12,5, 11, 10 und 6 kDa) identifiziert werden. Die Teilsequenz der Untereinheit 2 entspricht der aus der Literatur (Tzagoloff, A., 1982, Mitochondria, 111-130, Plenum Press New York) bekannten Sequenz.



**B. Konstruktion der scFv- oder Fab-Bank**

Die Sequenzen der variablen Region der leichten und schweren Ketten der Antikörper wurden genetisch über eine Linkersequenz, die  $(\text{Gly}_4\text{Ser})_{3-5}$  codiert, fusioniert (nach Clackson T. et al, Nature 352, 624-628, 1991; Ge, L. et al., in C.A.K. Borrebaeck (Hrsg.) Antibody Engineering 2. Aufl., Oxford University Press, New York, 1994). Alternativ kann auch eine menschliche Antikörperbank aus nicht immunisierten Spendern (vgl. Barbas III, a.a.O.) oder Consensussequenzen (EP-A-951 130 21) konstruiert werden. Ferner kann auch eine scFv-Bank, die sich von einer einzelnen Sequenz mit randomisiertem CDRH3 gemäß Barbas III et al., Gene, 137: 57-62, 1993 ableitet, konstruiert werden. CDRH3 kann unter Verwendung von NNK (N = A,C,G,T in gleichen molaren Verhältnissen, K = G und C) oder mit Codon-bezogener Mutagenese (vgl. US-A-5 264 563; Virnekäs, B., et al., Nuc. Acids Res., 22,5600-5607,1994) randomisiert werden.

Die Fab-Banken wurden gemäß Huse et al., Science 246: 1275-1281, 1989 konstruiert.

**C. Selektion von Antikörpern gegen die getrennten Proteine**

Ein Aliquot jedes der fraktionierten Proteine von Beispiel 1 wurde auf eine PVDF-Membran geblottet und zur Identifizierung markiert. Die Membran wurde mit 3% fettfreiem Milchpulver und UV-inaktivierten M13-Phagen blockiert. Eine auf einem filamentösen Phagen präsentierte scFv- oder Fab-Bank wurde mit der blockierten, die verschiedenen fraktionierten Proteine enthaltenden Membran

- 24 -

eine Stunde vermischt. Dann wurde die Membran ausgiebig gewaschen. Jedes der markierten Proteine wurde aus der Membran ausgeschnitten, und die gebundenen Phagen wurden mit 0,1 M TEA oder HCl 10 min eluiert und neutralisiert. Alternativ wurden die ausgeschnittenen Membranstücke mit PCR-Puffer getränkt, und das Eluat wurde als Matrize für die PCR-Reaktion unter Verwendung eines flankierenden Primerpaares, das die Amplifikation der scFv- oder Fab-Gene erlaubte, verwendet. Die Spezifität der auf der Oberfläche des gebundenen Phagen präsentierten scFvs oder Fabs wurde mittels ELISA oder Western Blotting bestimmt. Alternativ wurden die amplifizierten PCR-Fragmente in einen Expressionsvektor subcloniert und der Rohextrakt auf Bindung mittels ELISA oder Western Blotting getestet.

## **Beispiel 2**

### **A. Identifizierung der mitochondrialen Proteine aus dem 2D-Gel mittels Massenspektrometrie**

Die Proben aus fraktionierten mitochondrialen Proteinen wurden durch 2D-Gelelektrophorese getrennt. Das Gel wurde mit Coomassie angefärbt, und die Proteinflecken wurden mittels MALDI und Nanospray wie in Beispiel 1 beschrieben identifiziert.

Einer der Proteinflecken wurde als Rindermitochondrienprodukt ND2, eine Komponente der NADH-Dehydrogenase, identifiziert. Die N-terminale Sequenz stimmt mit der entsprechenden Sequenz aus der Literatur überein (Fearnley, und Walker,

a.a.O.).

Wie in Fearnley und Walker gezeigt, codiert das Codon ATG, ein universelles Isoleucincodon, in Rinderherzmitochondrien sowohl bei der Initiations- als auch bei der Elongationsstufe Methionin. Diese Information kann nur über die direkte Sequenzierung des Gens und Proteins erhalten werden.

#### **B. Selektion von Antikörpern gegen die mit dem 2D-Gel getrennten Proteine**

Die 2D-Elektrophoresegele wurden auf eine PVDF-Membran geblottet und mit einer Phagen-Antikörperbank wie in Beispiel 2 beschrieben abgesucht. Die durch irreversible Anfärbung des entsprechenden 2D-Gels markierten Proteinflecken wurden ausgeschnitten und die gebundenen Phagenteilchen wurden wie vorstehend beschrieben eluiert.

### **Beispiel 3**

#### **A. Konstruktion von Peptidbanken**

Peptidbanken wurden nach an sich bekannten Verfahren konstruiert (vgl. z.B. Devlin, J.J. et al., Science, 249. 404-406, 1990). Im Gegensatz zur scFv- oder Fab-Phagenbank können die präsentierten Peptide genetisch entweder mit dem Gen III (gIII), dem Minorhüllprotein von filamentösen Phagen, das für das Andocken des Phagen an den f-Pilus von E. coli und das Durchdringen der Wirtsmembran verantwortlich ist, oder dem Gen VIII (gVIII), dem Haupthüllprotein, fusioniert werden. Da pro Phagenteilchen nur 3-5 Kopien von gIIIp im Gegensatz

- 26 -

zu 2-3000 Kopien von gVIIIp vorhanden sind, können mehr Kopien des gleichen Peptids über eine gVIII-Fusion als über eine gIII-Fusion präsentiert werden, sind die ausgewählten Peptide als Folge des Einfangeffekts bei Präsentation auf gIII eher hochaffine Varianten als bei Präsentation auf gVIII.

#### **B. Selektion der Peptide gegen die getrennten Proteine**

Die fraktionierten Proben der mitochondrialen Proteine aus Beispiel 1 wurden mittels 2D-Gelelektrophorese oder Säulenchromatographie getrennt. Die fraktionierten Proteine wurden wie beschrieben auf eine Membran geblottet. Die Membran wurde blockiert und die Phagen-Peptid-Bank wurde direkt zu der Membran gegeben. Die für individuelle Proteine spezifischen Peptide wurden wie in Beispiel 1 beschrieben isoliert.

#### **Beispiel 4**

##### **A. Konstruktion einer Rinderherz-cDNA-Bank**

Eine Rinderherz-cDNA-Bank wurde von Stratagene (Kat. # 937722) bezogen und unter Verwendung des SurfZAP-Vektors (Stratagene) präsentiert.

##### **B. Selektion von cDNA gegen die getrennten Proteine**

Die fraktionierten Proben der mitochondrialen Proteine aus Beispiel 1 wurden mittels 2D-Gelelektrophorese oder Säulenchromatographie getrennt. Die fraktionierten Proteine wurden wie beschrieben auf eine Membran geblottet. Die

- 27 -

Membran wurde blockiert und die Phagen-cDNA-Bank wurde direkt zu der Membran gegeben. Die für individuelle Proteine spezifische cDNA wurden wie in Beispiel 1 beschrieben isoliert.

#### **Beispiel 5**

##### **A. Erzeugung kombinatorischer Rinderherzprotein-spezifischer Antikörperbanken**

Rinderherzgewebe wird nach einem Fachmann bekannten Standardmethoden homogenisiert. Das Homogenat wird unter Verwendung einer 2D-Gelelektrophorese getrennt und eine kombinatorische Antikörper-Phagenbank wird mit dem getrennten Homogenat in Kontakt gebracht und proteinspezifische Subbanken werden wie in Beispiel 2 isoliert.

##### **B. Identifizierung der Rinderherzproteine**

Die auf dem Gel getrennten Proteinproben werden geblottet und jeder Proteinfleck wird unter Verwendung von Massenspektrometrie wie in Beispiel 1 identifiziert.

##### **C. Expressionsprofil der Rinderproteine unter Verwendung der immobilisierten Proteinproben**

Das Rinderherzhomogenat wird auf Mikrotiterplatten immobilisiert. Die Anzahl der Kavitäten entspricht dabei der in Stufe A erzeugten Subbanken. Nach

- 28 -

Blockierung der Kavitäten mit geeigneten Puffern werden die Antikörpersubbanken zugesetzt und so inkubiert, daß ein Teil jeder Bank an das Homogenat in jeder Kavität bindet. Die Kavitäten werden gewaschen und die gebundenen Antikörper werden durch die an die Antikörper fusionierten Peptidtags identifiziert.

Es ist offensichtlich, daß mit geeigneten Modifikationen Proteinproben jeder biologischen Quelle auf jeder festen Oberfläche immobilisiert werden können und mit Antikörper- oder Peptidbanken nachgewiesen werden können. Umgekehrt können auch Antikörper- und Peptidbanken auf festen Oberflächen immobilisiert werden, und die Proteinproben werden mit der Oberfläche in Kontakt gebracht und durch physikalisch-chemische Verfahren wie z.B. Chemilumineszenz nachgewiesen.

**Patentansprüche**

1. Verfahren zur gleichzeitigen Identifizierung eines Proteinstems eines und seines Bindungspartners, dadurch gekennzeichnet, daß man
  - a) Proteine oder Proteinaggregate aus einer biologischen Quelle isoliert und auftrennt,
  - b) die aufgetrennten Proteine oder Proteinaggregate auf einer Oberfläche immobilisiert,
  - c) eine kombinatorische Bank mit den auf einer Oberfläche immobilisierten Proteinen oder Proteinaggregaten inkubiert,
  - d) diejenigen Mitglieder der kombinatorischen Bank, die an die immobilisierten Proteine binden, von nicht-gebundenen Mitgliedern der Bank trennt,
  - e) die an die Oberfläche gebundenen Komplexe aus Protein und Bindungspartner aus der kombinatorischen Bank isoliert,
  - f) die Proteine in den so isolierten Komplexen mit einer Kombination aus einem physikalisch-chemischen Verfahren identifiziert, und
  - g) gegebenenfalls die isolierten Bindungspartner anreichert.
2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß die biologische Quelle ein Organpräparat, eine Biopsieprobe, eine Körperflüssigkeit eines Lebewesens, eine Zellkultur, ein Mikroorganismus oder ein Pflanzenpräparat ist.
3. Verfahren nach einem der Ansprüche 1 oder 2, dadurch gekennzeichnet,

- 30 -

daß man bei Schritt a) die biologische Probe in einem geeigneten Puffer solubilisiert und die Probe unter Verwendung eines Proteintrennverfahrens ausgewählt aus 2D-Gelelektrophorese, Perfusionschromatographie, Flüssigchromatographie oder Kapillarelektrophorese auftrennt.

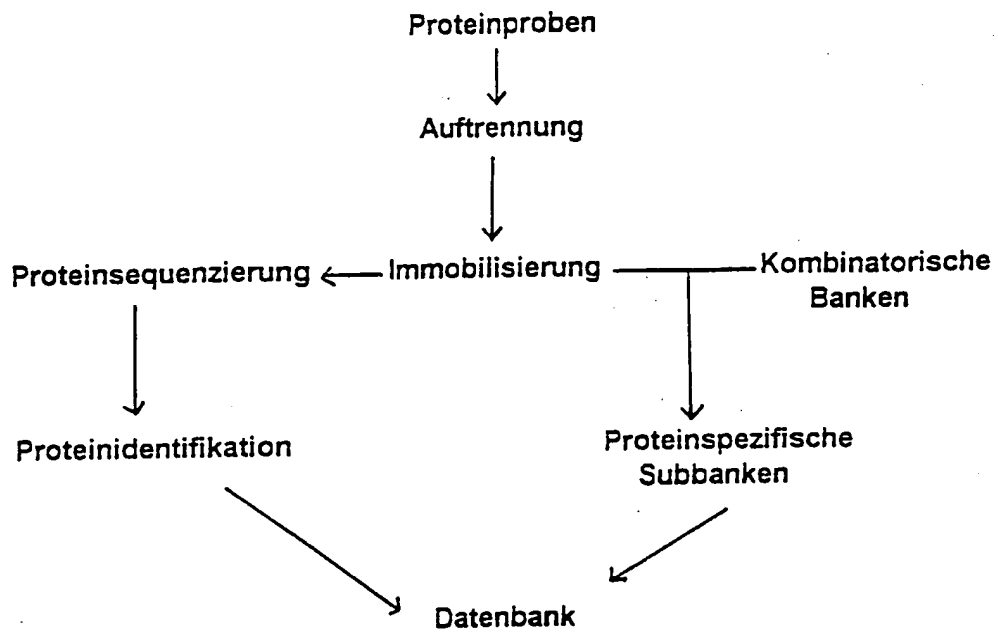
4. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, daß man in Schritt b) die aufgetrennten Proteine auf den Kavitäten einer Mikrotiterplatte immobilisiert oder auf eine Membran blottet oder mit spezifischen Antikörpern beschichteten Mikrokügelchen einfängt.
5. Verfahren nach einem der Ansprüche 1 bis 4, dadurch gekennzeichnet, daß man als kombinatorische Bank Random-Peptid-Banken, Banken der Immunglobulin-superfamilie, Protein-Display-Banken, kombinatorische chemische Banken, RNA- oder DNA-Banken verwendet.
6. Verfahren nach einem der Ansprüche 1 bis 5, dadurch gekennzeichnet, daß man bei Schritt d) die gebundenen Mitglieder der Bank direkt eluiert oder die in die Phagenpartikel, die die gebundenen Proteine präsentieren, gepackte DNA mit der PCR amplifiziert.
7. Verfahren nach einem der Ansprüche 1 bis 6, dadurch gekennzeichnet, daß man in Schritt g) einen bakteriellen Wirt mit den isolierten Phagen infiziert, um ausgewählte Phagenteilchen zu vermehren, und die ausgewählten Proteinbindungspartner sequenziert, oder die ausgewählten Bindungspartner durch einzigartige Sequenzanknüpfungen (Tag) identifiziert.



- 31 -

8. Verfahren nach einem der Ansprüche 1 bis 7, dadurch gekennzeichnet, daß man die aufgetrennten Proteine oder Proteinaggregate mittels Massenspektrometrie oder durch direktes Sequenzieren identifiziert.
9. Datenbank, umfassend Proteine und ihre spezifischen Bindungspartner und erhalten nach einem Verfahren nach einem der Ansprüche 1 bis 8.
10. Datenbank nach Anspruch 9 zur Verwendung zur gezielten Diagnose von Stoffwechselerkrankungen.
11. Verwendung einer Datenbank nach Anspruch 9 zur Entwicklung von Arzneimitteln, dadurch gekennzeichnet, daß man mit einem Verfahren nach einem der Ansprüche 1 bis 8 die Bindung einer Arzneimittelmodellverbindung an die Proteine eines Krankheitserregers ermittelt und die so erhaltenen Bindungsdaten qualitativ auswertet.
12. Diagnosechip, dadurch gekennzeichnet, daß er eine nach einem Verfahren nach einem der Ansprüche 1 bis 8 erhaltene proteinspezifische Bank oder Subbank eines menschlichen, tierischen oder pflanzlichen Gewebes immobilisiert an eine Oberfläche enthält.
13. Kit, dadurch gekennzeichnet, daß es einen Diagnosechip nach Anspruch 12 zusammen mit einer Probenahmeverrichtung, einem Reaktionsgefäß, Kalibrationskurven und Gebrauchsanweisung enthält.

1/1



Figur 1